CLAIMS

1000. A method of focused crawling, comprising:

accessing a query input, the query input including at least a first query part and a second query part;

5 crawling a plurality of documents, at least some of the plurality of documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the first query part; and

returning target documents, the target documents being relevant to the second query part, the target documents found from the plurality of crawled documents, the target 10 documents returned at least partly based on a search metric, the search metric at least partly determined by the mechanism and by the second query part.

1100. The method of claim 1000, wherein relevance includes importance.

18000. A method of focused crawling, comprising:

accessing a query input including at least a first query part and a second query part;

15 crawling a plurality of documents, at least some of the plurality of documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism and by the first query part; and

returning target documents, the target documents being relevant to the second query part, the target documents found from the plurality of crawled documents, the target 20 documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism and by the second query part.

18100. The method of claim 18000, wherein relevance includes importance.

2000. A method of focused crawling, comprising:

25 accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the

first combination including a first plurality of one or more procedures, the first plurality of one or more procedures including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the

5   template portions including a first plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents

10   returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including one or more of: 1) evaluating relevance of documents

15   using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents.

20   2200.   The method of claim 2000, wherein relevance includes importance.

2300.   The method of claim 2000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

25   2400.   The method of claim 2000, wherein one or more of: 1) the first plurality of one or more hierarchical levels and 2) the second plurality of one or more hierarchical levels, includes at least one or more heading levels and one or more content levels.

2500.   The method of claim 2000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more

30   referring documents and a second plurality of one or more referring documents, each of

the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

41000. A method of focused crawling, comprising:

5   accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of

10   one or more procedures including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4)

15   evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second

20   combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the

25   template portions including a second plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents,

wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents,

30   includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

5 assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the

10 ranked list at least partly generated from the graph.

41200. The method of claim 41000, wherein relevance includes importance.

41300. The method of claim 41000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

15 combining the evaluated relevances and the weights of the evaluated relevances.

41400. The method of claim 41000, wherein one or more of: 1) the first plurality of one or more hierarchical levels and 2) the second plurality of one or more hierarchical levels, includes at least one or more heading levels and one or more content levels.

41500. The method of claim 41000, wherein evaluating relevance includes evaluating

20 relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

25 41600. The method of claim 41000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

5   41700. The method of claim 41000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and

10   the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents

15   connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

41800. The method of claim 41000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the

20   crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all

25   documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first

30   plurality of documents.

41900. The method of claim 41000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

41a00. The method of claim 41000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

41b00. The method of claim 41000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

41c50. The method of claim 41b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

41d00. The method of claim 41000, wherein the propagating weights through the graph occurs up to a limited node distance.

41e00. The method of claim 41000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

41g00. The method of claim 41000, wherein relevance includes importance.

41i00. The method of claim 41000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

19000. A method of focused crawling, comprising:

accessing a query input, the query input being relevant to at least one of target documents and referring documents, each of the referring documents refer to at least one of the target documents;

crawling a plurality of documents, the documents including links to each other, and

5   the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly

10   determined by a second mechanism, the second mechanism being different from the first mechanism.

19200. The method of claim 19000, wherein relevance includes importance.

19300. The method of claim 19000, wherein at least one of the first mechanism and the second mechanism includes:

15   associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

19500. The method of claim 19000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

20   20000. A method of focused crawling, comprising:

accessing a query input; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the query input, the mechanism including a

25   combination, the combination including evaluating relevance of documents using a freshness of documents, and one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

20a00. The method of claim 20000, wherein the combination includes evaluating relevance of documents using a freshness of documents, and one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels, and 3) evaluating relevance of documents using a link structure of the crawled documents .

20100. The method of claim 20000, wherein relevance includes importance.

20400. The method of claim 20000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

20500. The method of claim 20000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

3000. A method of focused crawling, comprising:

accessing a query input; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the query input, the mechanism including a combination, the combination including evaluating relevance of documents using a link structure of the crawled documents, and one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

3a00. The method of claim 3000, wherein the combination including evaluating relevance of documents using a link structure of the crawled documents, and evaluating relevance based on freshness of documents, and one or more of: 1) evaluating relevance of

documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using content located in a specified part of a format.

3100. The method of claim 3000, wherein relevance includes importance.

3400. The method of claim 3000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

3500. The method of claim 3000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

42000. A method of focused crawling, comprising:

accessing a query input; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the query input, the mechanism including a combination, the combination including evaluating relevance of documents using a link structure of the crawled documents, and one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels,

wherein evaluating relevance of documents using a link structure of the crawled documents includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the
5   ranked list at least partly generated from the graph.

42100. The method of claim 42000, wherein the combination including evaluating relevance of documents using a link structure of the crawled documents, and evaluating relevance based on freshness of documents, and one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of
10   documents using content located in a specified part of a format.

42200. The method of claim 42000, wherein relevance includes importance.

42300. The method of claim 42000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

42400. The method of claim 42000, wherein evaluating relevance includes evaluating
15   relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

20   42500. The method of claim 42000, wherein evaluating relevance of documents using a link structure of the crawled documents further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the
25   plurality of received documents.

42600. The method of claim 42000, wherein evaluating relevance of documents using a link structure of the crawled documents further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and

the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

42700. The method of claim 42000, wherein evaluating relevance of documents using a link structure of the crawled documents further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

42800. The method of claim 42000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

42900. The method of claim 42000, wherein evaluating relevance of documents using a link structure of the crawled documents further comprises:

shrinking the graph by removing one or more nodes of the graph.

42a00. The method of claim 42000, wherein evaluating relevance of documents using a link structure of the crawled documents further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

42b50. The method of claim 42a00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

42c00. The method of claim 42000, wherein the propagating weights through the graph occurs up to a limited node distance.

42d00. The method of claim 42000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

42f00. The method of claim 42000, wherein relevance includes importance.

42h00. The method of claim 42000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

42i00. The method of claim 42000, further comprising:

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism and the query input.

43000. A method of focused crawling, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of one or more procedures including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3)

evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including evaluating relevance of documents using a template, the template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels,

wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

43200. The method of claim 43000, wherein relevance includes importance.

43300. The method of claim 43000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

43400. The method of claim 43000, wherein one or more of: 1) the first plurality of one or more hierarchical levels and 2) the second plurality of one or more hierarchical levels, includes at least one or more heading levels and one or more content levels.

43500. The method of claim 43000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

43600. The method of claim 43000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

43700. The method of claim 43000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

43800. The method of claim 43000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the

5   database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being

10   forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

43900. The method of claim 43000, wherein the first plurality of documents includes

15   recently received documents of the plurality of received documents.

43a00. The method of claim 43000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

20   43b00. The method of claim 43000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

25   43c50. The method of claim 43b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

43d00. The method of claim 43000, wherein the propagating weights through the graph occurs up to a limited node distance.

43e00. The method of claim 43000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

43i00. The method of claim 43000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

43j00. The method of claim 43000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a link structure of the crawled documents, and 3) evaluating relevance based on freshness of documents.

44000. A method of focused crawling, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of one or more procedures including evaluating relevance of documents using a link structure of the crawled documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second

plurality of procedures including evaluating relevance of documents using a template, the template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels,

wherein the procedure, of the first plurality of one or more procedures, of
5   evaluating relevance of documents using a link structure of the crawled documents, includes:

· accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

10   generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

15   generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

44200. The method of claim 44000, wherein relevance includes importance.

44300. The method of claim 44000, wherein at least one of the first mechanism and the second mechanism includes:

20   associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

44400. The method of claim 44000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

44500. The method of claim 44000, wherein evaluating relevance includes evaluating
25   relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document

directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

44600. The method of claim 44000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the
5   crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

10   44700. The method of claim 44000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and
15   the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents
20   connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

44800. The method of claim 44000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the
25   crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all
30   documents connected within a first specified number of links in a forward direction from

one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

44900. The method of claim 44000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

44a00. The method of claim 44000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

44b00. The method of claim 44000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

44c50. The method of claim 44b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

44d00. The method of claim 44000, wherein the propagating weights through the graph occurs up to a limited node distance.

44e00. The method of claim 44000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

44i00. The method of claim 44000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

44j00. The method of claim 44000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a link structure of the crawled documents, and 3) evaluating relevance based on freshness of documents.

44k00. The method of claim 44000, wherein the first plurality of one or more procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels, and 3) evaluating relevance based on freshness of documents.

45000. A method of focused crawling, comprising:

    accessing a query input;

    crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of one or more procedures including 1) evaluating relevance of documents using a link structure of the crawled documents and 2) evaluating relevance based on freshness of documents; and

    returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including evaluating relevance of documents using a template, the template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels,

wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

45200. The method of claim 45000, wherein relevance includes importance.

45300. The method of claim 45000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

45400. The method of claim 45000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

45500. The method of claim 45000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

45600. The method of claim 45000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

5   expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

45700. The method of claim 45000, wherein the procedure, of the first plurality of one or
10   more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

       expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the
15   plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or
20   more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

45800. The method of claim 45000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

25   expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from
30   one or more documents of the first plurality of documents, the forward direction being

forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

5   45900. The method of claim 45000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

45a00. The method of claim 45000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the

10   crawled documents, further comprises:

    shrinking the graph by removing one or more nodes of the graph.

45b00. The method of claim 45000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

15   shrinking the graph by combining one or more sets of one or more nodes of the graph.

45c50. The method of claim 45b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

45d00. The method of claim 45000, wherein the propagating weights through the graph

20   occurs up to a limited node distance.

45e00. The method of claim 45000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

45i00. The method of claim 45000, wherein evaluating relevance includes evaluating

25   relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

45j00. The method of claim 45000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a link structure of the crawled documents, and 3) evaluating relevance based on freshness of documents.

5   45k00. The method of claim 45000, wherein the first plurality of one or more procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

10   46000. A method of focused crawling, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the

15   first combination including a first plurality of one or more procedures, the first plurality of one or more procedures including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3)

20   evaluating relevance of documents using a first link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly

25   determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including 1) evaluating relevance of documents using a template, the template including a plurality of one or more template portions, at least one of the

template portions including a second plurality of one or more hierarchical levels, and 2) evaluating relevance of documents using a second link structure of the crawled documents,

wherein one or more of 1) evaluating relevance of documents using a first link structure of the crawled documents, and 2) evaluating relevance of documents using a

5   second link structure of the crawled documents includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

10   assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the

15   ranked list at least partly generated from the graph.

46200. The method of claim 46000, wherein relevance includes importance.

46300. The method of claim 46000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

20   combining the evaluated relevances and the weights of the evaluated relevances.

46400. The method of claim 46000, wherein one or more of: 1) the first plurality of one or more hierarchical levels and 2) the second plurality of one or more hierarchical levels, includes at least one or more heading levels and one or more content levels.

46500. The method of claim 46000, wherein evaluating relevance includes evaluating

25   relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document

directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

46600. The method of claim 46000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

46700. The method of claim 46000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

46800. The method of claim 46000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from

one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

46900. The method of claim 46000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

46a00. The method of claim 46000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

46b00. The method of claim 46000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

46c50. The method of claim 46b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

46d00. The method of claim 46000, wherein the propagating weights through the graph occurs up to a limited node distance.

46e00. The method of claim 46000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

46i00. The method of claim 46000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

46j00. The method of claim 46000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, and 2) evaluating relevance based on freshness of documents.

47000. A method of focused crawling, comprising:

5   accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of

10   one or more procedures including evaluating relevance of documents using a first link structure of the crawled documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly

15   determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including 1) evaluating relevance of documents using a template, the template including a plurality of one or more template portions, at least one of the

20   template portions including a plurality of one or more hierarchical levels, and 2) evaluating relevance of documents using a second link structure of the crawled documents,

wherein one or more of 1) evaluating relevance of documents using a first link structure of the crawled documents, and 2) evaluating relevance of documents using a second link structure of the crawled documents includes:

25   accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

47200. The method of claim 47000, wherein relevance includes importance.

47300. The method of claim 47000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

47400. The method of claim 47000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

47500. The method of claim 47000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

47600. The method of claim 47000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

47700. The method of claim 47000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

47800. The method of claim 47000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

47900. The method of claim 47000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

47a00. The method of claim 47000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

47b00. The method of claim 47000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

47c50. The method of claim 47b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

47d00. The method of claim 47000, wherein the propagating weights through the graph occurs up to a limited node distance.

47e00. The method of claim 47000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

47i00. The method of claim 47000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

47j00. The method of claim 47000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, and 2) evaluating relevance based on freshness of documents.

47k00. The method of claim 47000, wherein the first plurality of one or more procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a

template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels, and 3) evaluating relevance based on freshness of documents.

48000. A method of focused crawling, comprising:

5   accessing a query input;

   crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including a first plurality of one or more procedures, the first plurality of

10 one or more procedures including 1) evaluating relevance of documents using a first link structure of the crawled documents, and 2) evaluating relevance based on freshness of documents; and

   returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents

15 returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including a second plurality of one or more procedures, the second plurality of procedures including 1) evaluating relevance of documents using a template,

20 the template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels, and 2) evaluating relevance of documents using a second link structure of the crawled documents,

   wherein one or more of 1) evaluating relevance of documents using a first link structure of the crawled documents, and 2) evaluating relevance of documents using a

25 second link structure of the crawled documents includes:

   accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

   generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

48200. The method of claim 48000, wherein relevance includes importance.

48300. The method of claim 48000, wherein at least one of the first mechanism and the second mechanism includes:

associating a weight to each of the evaluated relevances of the procedures; and

combining the evaluated relevances and the weights of the evaluated relevances.

48400. The method of claim 48000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

48500. The method of claim 48000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

48600. The method of claim 48000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

48700. The method of claim 48000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

48800. The method of claim 48000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

48900. The method of claim 48000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

48a00. The method of claim 48000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by removing one or more nodes of the graph.

48b00. The method of claim 48000, wherein the procedure, of the first plurality of one or more procedures, of evaluating relevance of documents using a link structure of the crawled documents, further comprises:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

48c50. The method of claim 48b00, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

48d00. The method of claim 48000, wherein the propagating weights through the graph occurs up to a limited node distance.

48e00. The method of claim 48000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

48i00. The method of claim 48000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

48j00. The method of claim 48000, wherein the second plurality of procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, and 2) evaluating relevance based on freshness of documents.

48k00. The method of claim 48000, wherein the first plurality of one or more procedures further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a

template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

4000. A method of ranking documents, comprising:

5   accessing a plurality of documents, the documents including links to each other;

generating a graph representation of the plurality of documents, such that nodes of the graph represent the documents, and edges of the graph represent links linking the documents;

assigning weights to one or more nodes of the graph;

10   finding an approximate assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the approximate assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of the plurality of documents, the ranked list at least partly

15 generated from the graph.

4400. The method of claim 4000, wherein relevance includes importance.

21000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

20   generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents;

25   assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

21100. The method of claim 21000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

21200. The method of claim 21000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

21250. The method of claim 21200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

21300. The method of claim 21000, wherein the propagating weights through the graph occurs up to a limited node distance.

21400. The method of claim 21000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

22000. A method of ranking document comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

22100. The method of claim 22000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

22200. The method of claim 22000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

22250. The method of claim 22200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

22300. The method of claim 22000, wherein the propagating weights through the graph occurs up to a limited node distance.

22400. The method of claim 22000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

23000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or

more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

23100. The method of claim 23000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

23200. The method of claim 23000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

23250. The method of claim 23200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

23300. The method of claim 23000, wherein the propagating weights through the graph occurs up to a limited node distance.

23400. The method of claim 23000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

24000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and

the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being

5   forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

assigning weights to one or more nodes of the graph;

10   finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

15

24100. The method of claim 24000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

24200. The method of claim 24000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the

20   graph.

24250. The method of claim 24200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

24300. The method of claim 24000, wherein the propagating weights through the graph occurs up to a limited node distance.

25   24400. The method of claim 24000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

25000.   A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

5 expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents;

assigning weights to one or more nodes of the graph;

10 finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

15 25100. The method of claim 25000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

25200. The method of claim 25000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

20 25250. The method of claim 25200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

25300. The method of claim 25000, wherein the propagating weights through the graph occurs up to a limited node distance.

25400. The method of claim 25000, wherein weights assigned to a node include at least

25 one of relevance of the document to a query input and importance of the document independent of the query input.

25500. The method of claim 25000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

26000.   A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

5   generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

10   generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

26100. The method of claim 26000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

26200. The method of claim 26000, further comprising:

15   shrinking the graph by combining one or more sets of one or more nodes of the graph.

26250. The method of claim 26200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

26300. The method of claim 26000, wherein the propagating weights through the graph

20   occurs up to a limited node distance.

26400. The method of claim 26000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

26500. The method of claim 26000, wherein the first plurality of documents includes

25   documents of the plurality of received documents that are most recently received.

27000.   A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

5   expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction

10   from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

15   assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at

20   least partly generated from the graph.

27100. The method of claim 27000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

27200. The method of claim 27000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the

25   graph.

27250. The method of claim 27200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

27300. The method of claim 27000, wherein the propagating weights through the graph occurs up to a limited node distance.

27400. The method of claim 27000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

27500. The method of claim 27000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

28000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

28100. The method of claim 28000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

28200. The method of claim 28000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

28250. The method of claim 28200, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

28300. The method of claim 28000, wherein the propagating weights through the graph occurs up to a limited node distance.

28400. The method of claim 28000, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input.

28500. The method of claim 28000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

31000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of

weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

5

31250. The method of claim 31000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

32000. A method of ranking document, comprising:

10   accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

15   assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of

20   weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

32250. The method of claim 32000, wherein the combining is based on common

25   characteristics of the nodes or relationships between the nodes.

33000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the

5   plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or

10   more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a

15   node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights

20   propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

33250. The method of claim 33000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

25   34000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all

5   documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first

10   plurality of documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of

15   the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

20   generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

34250. The method of claim 34000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

35000. A method of ranking document, comprising:

25   accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents;

5   shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

10   finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at
15   least partly generated from the graph.

35250. The method of claim 35000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

35500. The method of claim 35000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

20   36000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

25   shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

36250. The method of claim 36000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

36500. The method of claim 36000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

37000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

37250. The method of claim 37000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

37500. The method of claim 37000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

38000. A method of ranking document, comprising:

accessing a first plurality of documents from a database of a plurality of received documents, the first plurality of documents to be ranked, the first plurality of documents including recently received documents of the plurality of received documents;

generating a graph of the first plurality of documents;

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of

the first plurality of documents, the backward direction being backward from the first plurality of documents;

shrinking the graph by removing one or more nodes of the graph and by combining one or more sets of one or more nodes of the graph;

assigning weights to one or more nodes of the graph, wherein weights assigned to a node include at least one of relevance of the document to a query input and importance of the document independent of the query input;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph up to a limited node distance, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

38250. The method of claim 38000, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

38500. The method of claim 38000, wherein the first plurality of documents includes documents of the plurality of received documents that are most recently received.

11000. A method of ranking documents, comprising:

accessing a plurality of documents, the documents including links to each other;

generating a graph representation of the plurality of documents, such that nodes of the graph represent the documents, and edges of the graph represent links linking the documents;

shrinking the graph by removing one or more nodes of the graph;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of the plurality of documents, the ranked list at least partly generated from the graph.

12000. A method of ranking documents, comprising:

5   accessing a plurality of documents, the documents including links to each other;

generating a graph representation of the plurality of documents, such that nodes of the graph represent the documents, and edges of the graph represent links linking the documents;

shrinking the graph by combining one or more sets of one or more nodes of the

10   graph;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

15   generating a ranked list of the plurality of documents, the ranked list at least partly generated from the graph.

13000. A method of ranking documents, comprising:

accessing a plurality of documents, the documents including links to each other;

20   generating a graph representation of the plurality of documents, such that nodes of the graph represent the documents, and edges of the graph represent links linking the documents;

assigning weights to one or more nodes of the graph;

propagating weights through the graph occurs up to a limited node distance;

25   finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

generating a ranked list of the plurality of documents, the ranked list at least partly generated from the graph.

5000. A method of focused crawling, comprising:

5 accessing input including one or more queries; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by at least one of a mechanism and at least part of the input, the mechanism at least evaluating relevance of documents using an approximate link structure of the crawled

10 documents.

5100. The method of claim 5000, wherein the link structure of the crawled documents is approximated by:

generating a graph representation of the plurality of documents, such that nodes of the graph represent the documents, and edges of the graph represent links linking the

15 documents;

assigning weights to one or more nodes of the graph;

finding an approximate assignment of weights to nodes of the graph, by propagating weights through the graph and calculating a weighted sum of weights propagated from neighboring nodes; and

20 generating a ranked list of the plurality of documents, the ranked list at least partly generated from the graph.

5200. The method of claim 5000, wherein the mechanism also includes one or more of:
1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using a template including a plurality of one or

25 more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

5300. The method of claim 5000, wherein relevance includes importance.

5500.   The method of claim 5000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

40000. A method of focused crawling, comprising:

5   accessing query input; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by at least one of a mechanism and at least part of the query input, the mechanism at least evaluating relevance of documents using a link structure of the crawled

10   documents,

wherein using the link structure includes:

accessing a first plurality of documents from a database of a plurality of received documents, the plurality of received documents including crawled documents, the first plurality of documents to be ranked;

15   generating a graph of the first plurality of documents;

assigning weights to one or more nodes of the graph;

finding an assignment of weights to one or more nodes of the graph, by propagating weights through the graph, the assignment of weight to a node based at least in part on calculating a weighted sum of weights propagated from neighboring nodes; and

20   generating a ranked list of at least the first plurality of documents, the ranked list at least partly generated from the graph.

40100. The method of claim 40000, wherein using the link structure further comprises:

expanding the graph with a second plurality of one or more documents from the database, wherein a third plurality includes a union of the first plurality of documents and

25   the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents.

40200. The method of claim 40000, wherein using the link structure further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) one or more documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) one or more documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

40300. The method of claim 40000, wherein using the link structure further comprises:

expanding the graph with a second plurality of one or more documents from the database, such that a third plurality includes a union of the first plurality of documents and the second plurality of documents, and the third plurality of documents is smaller than the plurality of received documents, the second plurality including one or more of: 1) all documents connected within a first specified number of links in a forward direction from one or more documents of the first plurality of documents, the forward direction being forward from the first plurality of documents, and 2) all documents connected within a second specified number of links in a backward direction from one or more documents of the first plurality of documents, the backward direction being backward from the first plurality of documents.

40400. The method of claim 40000, wherein the first plurality of documents includes recently received documents of the plurality of received documents.

40500. The method of claim 40000, further comprising:

shrinking the graph by removing one or more nodes of the graph.

40600. The method of claim 40000, further comprising:

shrinking the graph by combining one or more sets of one or more nodes of the graph.

40650. The method of claim 40600, wherein the combining is based on common characteristics of the nodes or relationships between the nodes.

40700. The method of claim 40000, wherein the propagating weights through the graph occurs up to a limited node distance.

40800. The method of claim 40000, wherein weights assigned to a document include at least one of relevance of the document to the query input and importance of the document independent of the query input.

40900. The method of claim 40000, wherein the mechanism also includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels.

40a00. The method of claim 40000, wherein relevance includes importance.

40c00. The method of claim 40000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

40d00. The method of claim 40000, further comprising:

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism and the query input.

6000. A method of finding starting points, comprising:

accessing a plurality of sample documents, the sample documents including links to other documents;

generating a graph representation of the plurality of sample documents and a plurality of referring documents, each of the referring documents referring to at least one

of the plurality of sample documents, such that nodes of the graph represent the sample documents and the referring documents, and edges of the graph represent links linking the sample documents and the referring documents; and

finding starting point documents, at least partly by performing a link structure analysis on the graph.

6100. The method of claim 6000, wherein the link structure analysis includes:

assigning weights at least to nodes representing sample documents; and

propagating weights through the graph in reverse direction , to nodes representing referring pages; and

assigning weights to nodes of the graph, by calculating a weighted sum of weights propagated from neighboring nodes.

6200. The method of claim 6000, wherein at least one of the plurality of referring documents refers to at least one of the plurality of sample documents directly.

6300. The method of claim 6000, wherein at least one of the plurality of referring documents refers to at least one of the plurality of sample documents indirectly through one or more documents.

6400. The method of claim 6000, wherein finding starting point documents further includes one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases and 2) evaluating relevance of documents using content located in a specified part of a format.

6600. The method of claim 6000, wherein relevance includes importance.

6700. The method of claim 6000, wherein the starting point documents provide starting points for at least one crawl.

6800. The method of claim 6000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

7000. A method of finding documents, comprising:

accessing a plurality of one or more documents;

accessing a search template, the search template includes a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels; and

returning target documents, the target documents found from the plurality of documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a mechanism, the mechanism includes at least evaluating relevance of documents to the search template.

7100. The method of claim 7000, wherein the accessed search template is determined from the plurality of sample documents.

7200. The method of claim 7000, wherein at least one of the heading level and the content level of at least one of the template potions is specified by at least a logical expression of keywords and phrases.

7300. The method of claim 7000, wherein at least one of the returned target documents has relevance to one or more template portions of the search template.

7400. The method of claim 7000, wherein the returned target documents are ranked at least partly based on relevance to a number of template portions of the search template.

7500. The method of claim 7000, wherein the accessed search template is retrieved from a repository of search templates.

7600. The method of claim 7000, wherein the accessed search template is constructed from the plurality of one or more documents.

7700. The method of claim 7000, wherein relevance includes importance.

7800. The method of claim 7000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

7900. The method of claim 7000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

8000. A method of focused crawling, comprising:

accessing at least a search template, the search template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels;

crawling a plurality of one or more documents, at least some of the plurality of one

5   or more documents including links to at least one other document of the plurality of one or more documents; and

returning target documents, the target documents found from the plurality of documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a mechanism, the mechanism includes at least

10   evaluating relevance of documents to the search template.

8100.   The method of claim 8000, wherein the accessed search template is determined from the plurality of sample documents.

8200.   The method of claim 8000, wherein at least one of the heading level and the content level of at least one of the template potions is specified by at least a logical

15   expression of keywords and phrases.

8300.   The method of claim 8000, wherein at least one of the returned target documents has relevance to one or more template portions of the search template.

8400.   The method of claim 8000, wherein the returned target documents are ranked at least partly based on relevance to a number of template portions of the search template.

20   8500.   The method of claim 8000, wherein the accessed search template is retrieved from a repository of search templates.

8600.   The method of claim 8000, wherein the accessed search template is constructed from the plurality of one or more documents.

8700. The method of claim 8000, wherein relevance includes importance.

25   8800. The method of claim 8000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

8900. The method of claim 8000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

9000. A method of focused crawling, comprising:

accessing a query input including at least a template, the template including a plurality of one or more template portions, at least one of the template portions including a plurality of one or more hierarchical levels; and

crawling a plurality of documents, the documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the query input, the mechanism including at least evaluating relevance of documents to the template.

9100. The method of claim 9000, wherein the accessed search template is determined from the crawled plurality of documents.

9200. The method of claim 9000, wherein at least one of the heading level and the content level of at least one of the template potions is specified by at least a logical expression of keywords and phrases.

9400. The method of claim 9000, wherein relevance includes importance.

9500. The method of claim 9000, wherein the plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

9600. The method of claim 9000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

10000. A method of building a repository, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other;

returning at least one of a plurality of one or more target documents and a plurality of one or more referring documents, the target documents returned being at last partly

relevant to the query input, the target documents found from the plurality of crawled documents; and

storing one or more of: the plurality of one or more target documents and the plurality of one or more referring documents.

10100. The method of claim 10000, wherein at least one of the plurality of referring documents refers to at least one of the plurality of target documents directly.

10200. The method of claim 10000, wherein at least one of the plurality of referring documents refers to at least one of the plurality of target documents indirectly through one or more documents.

10300. The method of claim 10000, further comprising:

accessing second query input;

searching the stored documents, at least partly responsive to the second query input; and

returning at least one of a second plurality of one or more target documents, the target documents returned being at last partly relevant to the second query input.

10400. The method of claim 10000, further comprising:

receiving a plurality of one or more fresh documents; and

updating one or more of the stored documents with plurality of one or more fresh documents.

10500. The method of claim 10400, wherein the plurality of one or more fresh documents is received if one or more of the stored documents has changed.

10600. The method of claim 10000, further comprising:

crawling a second plurality of documents from at least one of the stored documents.

10700. The method of claim 10000, wherein relevance includes importance.

14000. A system, comprising:

a first processor; and

a first plurality of one or more processors, wherein the first processor and each of the first plurality of one or more processors at least partly performs:

accessing a first query input and a second query input;

crawling a plurality of documents, at least some of the plurality of documents including links to each other, the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a mechanism and by the first query input; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by the mechanism and by the second query input.

14100. The system of claim 14000, wherein relevance includes importance.

15000. A method, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including one or more of: 1) evaluating relevance of documents using

logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating

5 relevance based on freshness of documents,

wherein the method is performed on at least one of 1)a first processor and 2) one or more of a first plurality of one or more processors.

15300. The method of claim 15000, wherein at least a first task of a plurality of tasks is performed on at least the first processor, and at least a second task of the plurality of tasks

10 is performed on at least one processor of the first plurality of one or more processors, and the plurality of tasks includes:

1) computing starting points from sample documents;

2) computing the link structure of the crawled documents;

3) evaluating relevance of documents using logical expressions of keywords and

15 phrases;

4) evaluating relevance of documents using the template; and

5) fetching the crawled documents.

15310. The method of claim 15300, wherein fetching the crawled documents includes fetching at least one document from at least one remote server.

20 15200. The method of claim 15000, wherein relevance includes importance.

15800. The method of claim 15000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document

25 directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

15500. The method of claim15000, wherein the first plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

15600. The method of claim 15000, wherein the second plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

15700. The method of claim 15000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

15400. The method of claim 15000, wherein at least one task of a plurality of tasks is performed on at least two processors, and the plurality of tasks includes:

    1) computing starting points from sample documents;

    2) computing the link structure of the crawled documents;

    3) evaluating relevance of documents using logical expressions of keywords and phrases;

    4) evaluating relevance of documents using the template; and

    5) fetching the crawled documents.

15410. The method of claim 15400, wherein fetching the crawled documents includes fetching at least one document from at least one remote server.

16000. A method, comprising:

    performing a plurality of focused crawls, wherein each of the plurality of focused crawls comprises:

        accessing a query input;

        crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly determined by a first mechanism, the first mechanism including a first combination, the first combination including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3) evaluating

relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents,

wherein the method is performed on at least one of 1)a first processor and 2) one or more of a first plurality of one or more processors.

16300. The method of claim 16000, wherein the plurality of focused crawls includes at least a first focused crawl and a second focused crawl, the first focused crawl performs a first plurality of tasks, the second focused crawl performs a second plurality of tasks,

wherein the first plurality of tasks includes:

1) computing starting points from sample documents;

2) computing the link structure of the crawled documents;

3) evaluating relevance of documents using logical expressions of keywords and phrases;

4) evaluating relevance of documents using the template; and

5) fetching the crawled documents, and

wherein the second plurality of tasks includes:

6) computing starting points from sample documents;

7) computing the link structure of the crawled documents;

8) evaluating relevance of documents using logical expressions of keywords and phrases;

9) evaluating relevance of documents using the template; and

10) fetching the crawled documents, and

wherein and at least one task of the first plurality of tasks is performed on at least the first processor, and at least one task of the second plurality of tasks is performed on at least one processor of the first plurality of processors.

16310. The method of claim 16300, wherein fetching the crawled documents includes fetching at least one document from at least one remote server.

16200. The method of claim 16000, wherein relevance includes importance.

16500. The method of claim 16000, wherein the first plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

16600. The method of claim 16000, wherein the second plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

16700. The method of claim 16000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

16800. The method of claim 16000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first document indirectly through one or more documents.

17000. A method, comprising:

accessing a query input;

crawling a plurality of documents, the documents including links to each other, and the crawling at least partly guided by a crawl metric, the crawl metric at least partly

determined by a first mechanism, the first mechanism including a first combination, the first combination including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a first plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents; and

returning target documents, the target documents being relevant to the query input, the target documents found from the plurality of crawled documents, the target documents returned at least partly based on a search metric, the search metric at least partly determined by a second mechanism, the second mechanism including a second combination, the second combination being different from the first combination, the second combination including one or more of: 1) evaluating relevance of documents using logical expressions of keywords and phrases, 2) evaluating relevance of documents using a template including a plurality of one or more template portions, at least one of the template portions including a second plurality of one or more hierarchical levels, 3) evaluating relevance of documents using a link structure of the crawled documents, and 4) evaluating relevance based on freshness of documents,

wherein, during at least part of a time interval, the time interval beginning at a first time and ending at a second time, the first time being a sending of an access request for a document from a remote server, the second time being a receiving of the document from the remote server, the method performs at least one of a plurality of tasks, the plurality of tasks including:

1) computing the link structure of the crawled documents;

2) evaluating relevance of documents using logical expressions of keywords and phrases;

3) evaluating relevance of documents using the template;

4) requesting a second document; and

5) receiving a third document.

17200. The method of claim 17000, relevance includes importance.

17500. The method of claim 17000, wherein the first plurality of one or more hierarchical levels includes at least one or more heading levels and one or more content levels.

17600. The method of claim 17000, wherein the second plurality of one or more

5   hierarchical levels includes at least one or more heading levels and one or more content levels.

17700. The method of claim 17000, wherein evaluating relevance of documents includes evaluating relevance of at least a first document and a second document, the second document referring to the first document.

10   17800. The method of claim 17000, wherein evaluating relevance includes evaluating relevance of at least a first document and one or more of a first plurality of one or more referring documents and a second plurality of one or more referring documents, each of the first plurality of one or more referring documents referring to the first document directly, and each of the second plurality of referring documents referring to the first

15   document indirectly through one or more documents.